WHAT IS CLAIMED IS:

1. A method of parsing text to form a logical representation of the text, the logical representation having tokens representing non-terminals and words of the text, the method comprising:

selecting a token;

identifying an integer that represents the selected token; and

utilizing the integer to identify at least one token of the logical representation that begins with the selected token.

2. The method of claim 1 wherein identifying an integer comprises identifying an integer that points to an identifier array, each cell in the identifier array providing a token identifier for a token that begins with the selected token.

3. The method of claim 2 wherein the token identifiers are integers.

4. The method of claim 3 wherein each token identifier integer comprises a table identifying portion and an offset portion, the table identifying portion identifying a table that contains an array of definitions for tokens including the token represented by the token identifier integer and the

offset portion identifying the location of the definition for the token represented by the token identifier integer.

5. The method of claim 4 wherein each cell in the identifier array further provides an indication of a rule in which the token represented by the token identifier integer begins with the selected token.

6. A computer-readable medium having a data structure used in parsing text, the data structure comprising:

a table of integers, each integer representing a token; and

a table of arrays, wherein each integer in the table of integers acts as a pointer to an array in the table of arrays and wherein each cell in the array contains a token identifier for a token activated by the token represented by the integer that points to the array.

7. The computer-readable medium of claim 6 wherein each token identifier comprises an integer.

8. The computer-readable medium of claim 7 wherein the token identifier points to a token definition for a token.

9. The computer-readable medium of claim 8 wherein the token definition for a token comprises a sequence of token identifiers that can be parsed to form the token defined by the token definition.

10. The computer-readable medium of claim 9 wherein each cell in the array further comprises a pointer to a sequence of tokens in the token definition.

11. A method of parsing text to form a tokenized representation of the text, the method comprising:

selecting a word from the text;

forming a partial parse of a token based on the selected word;

identifying an item that is needed to extend the partial parse; and

placing a pointer to the partial parse in a table associated with a next word in the text, the pointer being mapped from the item that is needed to extend the parse.

12. The method of claim 11 wherein placing a pointer comprises placing a pointer to an array in the table, wherein the array contains at least two partial parses that can be extended by a same item.

13. The method of claim 11 wherein identifying an item comprises identifying a word.

14. The method of claim 11 wherein identifying an item comprises identifying a non-terminal.

15. A computer-readable medium having a data structure used in parsing, the data structure comprising:

a set of mappings, each mapping associated with an item needed to extend a partial parse of a token; and

at least one pointer for each mapping, each pointer identifying at least one partial parse structure that needs the item associated with the mapping.

16. The computer-readable medium of claim 15 wherein at least one pointer points to an array identifying two different partial parse structures.

17. The computer-readable medium of claim 15 wherein the item needed is a word.

18. The computer-readable medium of claim 15 wherein the item needed is a non-terminal.

19. The computer-readable medium of claim 15 wherein each mapping maps from an integer that represents the item needed to extend a partial parse.

20. A method of parsing text to form a representation of the text, the representation having structures that span sub-strings of words in the text, each structure having a token at its root, the method comprising:

identifying a first structure that spans a first sub-string of words in the text and has a first token as its root, the first sub-string having a starting position and an ending position;

indexing the first structure by the first token and the staring position and ending position of the first sub-string;

identifying a second structure that spans the first sub-string of words and has the first token as its root;

using the first token and the start position and end position of the first sub-string to locate the first structure; and

removing one of the first structure and second structure from further consideration in the formation of the representation of the text.

21. The method of claim 20 wherein removing one of the first structure and second structure comprises removing the second structure.

22. The method of claim 20 wherein removing one of the first structure and second structure comprises removing the first structure.

23. The method of claim 22 wherein removing the first structure comprises removing the first structure so that it is no longer indexed by the first token and the starting position and ending position of the first sub-string and indexing the second structure by the first token and the starting position and ending position of the first sub-string.

24. The method of claim 20 wherein removing one of the first structure and the second structure comprises comparing the first structure to the second structure to determine which structure is better for the representation of the text.

25. A computer-readable medium having a data structure, the data structure comprising:

a first address field containing a token, the first address field for indexing an entry field that designates parse structures, the parse structures in an

entry field having the token of the first address field as their root node; and

a second address field for further indexing the entry field, the second address field containing a representation of words spanned by a parse structure.

26. The computer-readable medium of claim 25 wherein the second address field comprises a starting point and an ending point for a set of words in a text string.

27. A method of parsing text to identify a parse structure containing tokens that represent non-terminals and words, the method comprising:

converting a selected token into a token ID;

using a first portion of the token ID to identify a table containing definitions for tokens of a same type as the selected token;

using a second portion of the token ID to locate the definition for the selected token in the identified table; and

using the definition for the selected token as part of the method of identifying the parse structure.

28. A computer-readable medium having a data structure used in parsing text into a logical form containing tokens, the data structure comprising:

a collection of tables wherein each table is associated with a type of token and each table comprises a collection of definitions for tokens of the type; and

a set of token IDs, each token ID representing a separate token and comprising:

a first part that indicates a table that contains the definition for the token represented by the token ID; and

a second part that indicates the location of the definition of the token represented by the token ID within the table identified by the first part.